Principal Component Analysis (PCA)

It is one of the most intuitively simple and frequently used methods for applying dimension reduction in data and projecting it onto an orthogonal subspace of features.

In a very general form, it can be represented as the assumption that our observations look like some ellipsoid in the subspace of our original space.

Our new basis in this space coincides with the axes of this ellipsoid.

The dimension of this ellipsoid is equal to the dimension of the original space, but our assumption that the data lies in a subspace of a smaller dimension allows us to discard the other subspaces in the new projection.

One by one is OK: $\longrightarrow$ we could do it greedily

Our target:

to reduce the dimension of our data from n to k, $k \leq n$, we need to choose the top k axes of such an ellipsoid, sorted in descending order by dispersion along the axes.

$$\mathrm{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j$$

Here μ_i is the mean of the i^{th} feature.

covariance matrix

$$\Sigma = E[(X - E[X])(X - E[X])^T]$$

The covariance matrix is a generalization of variance in the case of multidimensional random variables

maximum variance of dataset $\longleftarrow$ / maximum eigenvalue /

eignvector $w_i \longleftrightarrow$ eigenvalue λ_i

$$\underset{n \times n}{\Sigma} \underset{n \times 1}{w_i} = \lambda_i \underset{n \times 1}{w_i}$$

If a single vector, than $\underset{1\times n}{v^T}\,\underset{n\times n}{X}$ is an array of projections of our observations.

multiple.. $\underset{i\times n}{V^T}\,\underset{n\times n}{X}$.

if $i < n$, then we lose some information.